



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Are single and extended metaphors processed differently? A test of two Relevance-Theoretic accounts

Citation for published version:

Rubio-Fernandez, P, Cummins, C & Tian, Y 2016, 'Are single and extended metaphors processed differently? A test of two Relevance-Theoretic accounts', *Journal of Pragmatics*, vol. 94, pp. 15-28.
<https://doi.org/10.1016/j.pragma.2016.01.005>

Digital Object Identifier (DOI):

[10.1016/j.pragma.2016.01.005](https://doi.org/10.1016/j.pragma.2016.01.005)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Journal of Pragmatics

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Are single and extended metaphors processed differently?

A test of two Relevance-Theoretic accounts

Abstract

Carston (2010) proposes that metaphors can be processed via two different routes. In line with the standard Relevance-Theoretic account of loose use, single metaphors are interpreted by a local pragmatic process of meaning adjustment, resulting in the construction of an ad hoc concept. In extended metaphorical passages, by contrast, the reader switches to a second processing mode because the various semantic associates in the passage are mutually reinforcing, which makes the literal meaning highly activated relative to possible meaning adjustments. In the second processing mode the literal meaning of the whole passage is metarepresented and entertained as an ‘imaginary world’ and the intended figurative implications are derived later in processing. The results of three experiments comparing the interpretation of the same target expressions across literal, single-metaphorical and extended-metaphorical contexts, using self-paced reading (Experiment 1), eye-tracking during natural reading (Experiment 2) and cued recall (Experiment 3), offered initial support to Carston’s distinction between the processing of single and extended metaphors. We end with a comparison between extended metaphors and allegories, and make a call for further theoretical and experimental work to increase our understanding of the similarities and differences between the interpretation and processing of different figurative uses, single and extended.

Introduction

Working within a Relevance Theory approach, Carston (2010) argues that some metaphors are analogous to other instances of loose use, and that for these cases the linguistic meaning of the metaphoric vehicle is pragmatically modulated during comprehension. That is to say, the truth-conditional content (explicature, in RT terms) of the metaphor (1) is that the lawyer in question is a SHARK*. The ad hoc concept SHARK* is a pragmatically adjusted version of the lexically-encoded concept SHARK, with its denotation including a certain type of human beings as well as the marine creatures.

(1) My lawyer is a shark.

This view of the truth-conditional content of metaphors agrees with that of Glucksberg and Keysar (1990), who contend that metaphors such as (1) are indeed class-inclusion assertions, just as they appear to be. The process of obtaining the classically “metaphorical” meanings of (1) – such as that the lawyer is ruthless, merciless, etc. – is then presumed to be an inferential pragmatic process that involves the enhancement of relevant properties of the metaphor vehicle (here, “shark”) and the suppression of irrelevant properties (in this case, IS A FISH, SWIMS IN THE SEA, HAS FINS, etc.), although the precise nature of this process is disputed (see also Sperber and Wilson 2002, Recanati 2004).

As Carston discusses, this kind of analysis is at odds with the intuition about metaphor articulated by Davidson (1978/1984: 263), who considers metaphors to have no propositional content other than their literal meaning and to be essentially imagistic, observing that “much of what we are caused [by metaphor] to notice is not propositional in character”. In Carston’s view, a comprehensive account of metaphor must acknowledge both the propositional and the imagistic component of metaphoric meaning, but also allow for the relative weights of these components to differ from case to case. By contrast with examples such as (1) in which the

propositional meaning predominates, Carston posits a second broad class of metaphors in which the literal meaning and its accompanying imagery is more dominant, with implicated propositional content only derived at a later stage (although possibly to some extent in parallel with the processing of the literal meaning). In particular, she analyses cases of extended metaphor such as (2) as belonging to this latter class.

- (2) All the world's a stage,
And all the men and women merely players.
They have their exits and their entrances
And one man in his time plays many parts

(Shakespeare, *As You Like It*, Act 2, Scene VII)

From the point of view of the ad hoc concept (propositional) approach, the interpretation of an extended metaphor would seem to require the construction of a series of ad hoc concepts: STAGE*, PLAYERS*, EXITS*, ENTRANCES* and PARTS* must all be constructed by the hearer, each reflecting the modulation of the appropriate literal meaning by pragmatic considerations. Carston (2010: 308) argues that instead, in such developed metaphor cases, the literal meaning of the words may take over from any adjustment process as the various strongly related literal meanings of the metaphor vehicles are mutually semantically primed and reinforced. So the literal meaning of the extended metaphor is entertained as a whole and it is this whole scenario that is the basis for the derivation of the ultimate 'metaphorical meaning'.

The expectation that the literal meaning of an extended metaphor should have this kind of internal coherence does not follow from the propositional account discussed above; that approach has little to say about how the literal meaning of one metaphor vehicle should influence the interpretation of subsequent metaphorically used words. Nevertheless, it seems clear that hearers do have these expectations, as shown by the perceived oddness of mixed

metaphors such as the following (Denys Parsons, letter in the New Scientist, 25 March 1971, p.704).

- (3) It is the thin end of a white elephant.
- (4) [They] have gone out of their way to rub in the bitter pill.
- (5) Mr Dewey would have been wielding a double-edged sword in the shape of a boomerang that would have come home to plague him.

In a conceptual modulation account, these extended metaphors should be unproblematic. For instance, in (5), the concepts of DOUBLE-EDGED SWORD* and BOOMERANG* are completely compatible: both refer to a means of attacking someone that could also injure the person who uses it. The incongruity seems to arise because of the expectation that the metaphorical expressions will be internally coherent at a literal level as well as a metaphorical level, and this fails because of the difficulty of imagining a “double-edged sword in the shape of a boomerang”. The problem of mixed metaphor could be taken to support the claim that the interpretation of extended metaphors has a literal-imagistic component in addition to the ultimate propositional component, as Carston contends.

Processing single and extended metaphors

From a processing standpoint, one way to interpret the distinction between single (lexical/phrasal) metaphors and extended-imagistic metaphors is as a claim about the time-course by which the metaphorical meanings are recovered by the hearer. Extended metaphors have to be internally coherent as regards their literal meaning (i.e. coherent within the imaginary world being described), but that is not all: they must, of course, also be interpretable as metaphors. The propositional account of metaphor interpretation would hold

that the appropriate ad hoc concepts (in the Shakespearian example, EXITS*, ENTRANCES*, etc.) are constructed on-line, and hence could be argued to predict that the metaphorical meanings associated with these ad hoc concepts (“death”, “birth”, etc.) should be immediately activated in the mind of the hearer. By contrast, the literal-imagistic account is not committed to this position: in extended metaphors “the literal meaning takes over and is metarepresented as a whole” (Carston 2010: 312) and the metaphoric meaning emerges more gradually from this holistic representation, in the form of an array of weakly implicated propositions (e.g., “human life is short”, “each individual’s actions are largely predictable and pre-determined”, “human existence has no greater meaning or purpose than a passing performance on a stage”, and so on).

Existing experimental research on metaphor has not directly addressed this question, although it has begun to probe the time-course of literal and figurative associates of metaphorical expressions. In a series of studies, Glucksberg, Gernsbacher and colleagues (see Glucksberg et al. 2001) explored the effect of metaphors on the subsequent truth-value judgement of literal sentences. They showed that prior exposure to a sentence such as (6) speeded up verification of (7) but slowed down verification of (8) compared to a baseline, which they interpreted as evidence that metaphor interpretation does indeed involve enhancing relevant properties and suppressing irrelevant properties of the metaphor vehicle.

(6) That defence lawyer is a shark.

(7) Sharks are tenacious.

(8) Sharks are good swimmers.

Augmenting these findings, Rubio-Fernández (2007) further explored the persistence of literal meaning in the interpretation of metaphor, using a cross-modal lexical priming paradigm. Analogously to classic research on the disambiguation of homonyms (Swinney

1979 and others), Rubio-Fernández tested the time-course over which exposure to a metaphor vehicle would prime literally related words. She showed that these were primed at 0ms and 400ms offsets from the word recognition point from the metaphor vehicle, but that priming at 1000ms offset was only present for metaphorically relevant words. That is, given the prime (9), the recognition of “cat” as a real word was primed at 0ms and 400ms offset from the word “cheetah” but not at 1000ms offset, whereas the recognition of “fast” was primed at all three offsets.

(9) Nobody wanted to run against John at school. John was a cheetah.

In sum, then, the evidence from studies of single metaphors suggests that properties that are relevant for the metaphorical understanding (in terms of the propositional account, those that are part of the ad hoc concept) are promoted by hearers and remain highly activated for at least a second. By contrast, properties that are literally associated with the metaphor vehicle but are not part of the metaphorical meaning (those that are not part of the ad hoc concept) are initially activated but are suppressed sometime after 400ms has elapsed.

The research discussed above, however, has concerned itself primarily with single metaphors, in which a single property is directly predicated of an entity (as in “John was a cheetah”). This confers certain advantages in making the materials easier to control and construct, as well as making it possible to compare metaphors with their minimally-different simile counterparts (a point of theoretical interest to researchers such as Glucksberg and Haught (2006)), but might diminish the generalisability of the results. Moreover, the way in which the materials were presented to participants could be argued to influence participants to compute the figurative meanings of metaphoric expressions as rapidly as possible. For instance, in Rubio-Fernández’s (2007) experiment, participants are explicitly instructed that each text ends in a metaphor, and that they should try to visualise the figurative meaning of

that metaphor (a request made in order to ensure that participants processed the stimuli as metaphors). Under these instructions it would be irrational for participants to switch to any literal-imagistic mode of processing, as it is purely the figurative meaning and not the literal meaning of the metaphor that the participant is instructed to recover. Literary texts, of course, do not come with any such instructions, so whether and to what extent readers persist in activating the literal associates remains an open question.

The present study

While Carston's (2010) account of metaphor interpretation is psychologically oriented, a number of empirical questions remain open. For example, at what point does the reader switch from a literal to a figurative mode of processing when reading extended metaphors? And more generally, how do the two processing modes interact? Another important question is whether the two processing modes work differently depending on the type of metaphors encountered in an extended passage. For example, while it is conceivable that the literal meanings of relatively novel metaphors prime each other during processing, conventional metaphors may be more strongly associated with their figurative meanings and their literal interpretation may be relatively less accessible. This, in turn, may prevent readers from switching to a literal-imagistic mode of processing. This potential exception to Carston's second processing route is related to a fundamental assumption of Lakoff and Johnson's (1980) view that people routinely use conceptual mappings to understand conventional expressions in discourse (a view that has been empirically tested with mixed results; see, e.g., Shen & Balaban, 1999; Keysar et al., 2000; Gentner et al. 2002; Thibodeau & Durgin, 2007). Given the open questions noted above, it is not straightforward to derive empirically testable predictions from Carston's (2010) account of metaphor interpretation. However, there are

fundamental differences between Carston's account and the standard Relevance-Theoretic account of loose use, which are potentially amenable to experimental investigation. The goal of this paper is therefore to explore whether readers of single and extended metaphors exhibit processing differences that could suggest, in line with Carston's (2010) account, that extended metaphors are first interpreted literally-imagistically before the global metaphorical meaning is derived off-line, whereas single metaphors are propositionally and interpreted on-line, as suggested by the traditional RT view.

We aim to test potential differences between the processing single and extended metaphors by comparing the comprehension of the same expression in three distinct contexts: an extended metaphorical context (where there are multiple figurative uses in the passage, including the target expression), a single-metaphorical context (where the target expression is the only figurative use in the passage) and a literal context (where the entire passage is literal). In all cases, the target expression appears in the second half of the passage. The following triplet is a sample item, with the target expression in bold (for more items see the Appendix).

Extended metaphor: The school's pupils were notoriously unruly. The place was a zoo. New keepers were generally assigned to the safer animals: the half-domesticated goats, the placid giraffes, the comical penguins. Those animals you could handle **without protective gloves**. It was both the more intelligent and the more aggressive animals that required extra safety measures.

Single metaphor: The school's pupils were notoriously unruly. New teachers were generally assigned to the less troublesome classes: the younger kids, the more settled ones, and the more relaxed ones. They seldom made a lot of trouble,

and you could work with them **without protective gloves**. It was both the more intelligent and the more aggressive pupils that required specialist care.

Literal control: The animals at the sanctuary had come from many different places. Some of them had been pets of owners who had died or had to move away. They were usually docile, and you could handle them **without protective gloves**. Other animals had been abused, or even trained to attack. They required special safety measures.

~~A possible criticism to this experimental design is that by manipulating the contexts preceding the target expressions, we do not only distinguish the different conditions but also introduce a large number of linguistic factors that could in principle affect the processing of the target expressions (e.g., varying word lengths and frequencies, different syntactic constructions, potential priming effects, or uneven flow of the narrative, just to name a few). While we acknowledge this shortcoming of our paradigm, we see it as inevitable insofar as it would be almost impossible to control for such a vast suite of potential factors in the construction of the materials. In other words, it would be extremely difficult to construct a battery of single metaphors, extended metaphors and literal controls that were preceded by contexts containing words of comparable length and frequency, identical syntactic constructions, the same potential priming effects, etc., while making sure that the passages are highly comprehensible.~~

~~Most important for the present investigation, the potential effects that linguistic factors might have on the processing of the target expressions do not result in any specific hypotheses that could be pinned against the two Relevance-Theoretic accounts of metaphor interpretation investigated in this study. Therefore, in our view, the best we can do when comparing the processing of single and extended metaphors is to acknowledge the variability inherent in the~~

~~materials, and spell out in detail the predictions that follow from Carston's (2010) model and the standard Relevance Theoretic account based wholly on ad hoc concepts (e.g., Sperber & Wilson 2008).~~

In the extended metaphorical contexts, Carston's account can naturally be taken to predict that the literal meaning of the target expressions should be accessed and maintained, on the basis that readers should be in a literal mode of processing at this stage (as the literal meaning is primed by various associates, e.g., "zoo", "goats", "giraffes", "penguins" and "animals" in the above example). By contrast, the single metaphors would be interpreted figuratively, by the reader pragmatically modulating the meaning of the metaphor vehicle, specifically broadening it radically to accommodate the metaphorical meaning. Consequently, Carston's account predicts that (a) initially patterns of comprehension should be similar in the literal control and extended metaphor conditions (as these both initially receive literal interpretations) and that (b) these should both contrast with the single metaphor condition (whose figurative interpretation is recovered immediately). These interpretive differences are compatible with more directly measurable differences in processing effort, with the construction of ad hoc concepts in single-metaphorical contexts potentially requiring more processing resources than the literal interpretation of the target expressions in literal and extended metaphorical contexts.

By contrast, on a standard RT account, both the single and extended metaphor conditions require the hearer to construct ad hoc concepts, and so would be expected to give rise to similar processing patterns. We might expect these to contrast with the literal condition, on the assumption that the hearer does not need to construct ad hoc concepts in the literal condition: the lexical concept will suffice, and no pragmatic modulation is required. Literal-first accounts of metaphor interpretation such as those of Searle (1979) or Grice (1975) would also predict a contrast between the literal control and the metaphorical conditions. However,

given that the RT account holds that pragmatic modulation of word meanings is widespread (cf. Wilson & Carston 2007), it would also be compatible with a view in which all three conditions required the hearer to construct ad hoc concepts and hence all gave rise to broadly identical processing signatures.

In summary, only Carston's account predicts processing differences between single and extended metaphors, as a consequence of these types of expression being interpreted via different routes (i.e. ad hoc concept/propositional vs. literal/imagistic respectively). By contrast, the standard RT account of metaphor interpretation (based solely on ad hoc concept construction), as well as more traditional metaphor accounts (based on a literal-figurative distinction) wouldn't predict processing differences between single and extended metaphors. While the absence of such differences would not necessarily disprove Carston's account, under those circumstances, the traditional RT account could be argued to be more parsimonious in that it posits a single mode of metaphor processing.

Pre-test of the materials

In order to examine the processing of extended and single metaphors, we needed to establish a set of novel metaphorical expressions. Specifically, we required items whose metaphorical meanings could successfully be recovered by participants, but which were sufficiently unfamiliar that there was no possibility of their meanings being literally represented (as in the case of "dead metaphors" such as 'the foot of a hill'). For this reason, we constructed and pre-tested our materials as discussed below.

Method

Participants

16 undergraduate and postgraduate students from University College London took part in the experiment. They were all native speakers of English and participated for monetary compensation (£3).

Materials, design and procedure

We constructed 21 extended metaphors (average number of words: 59; range: 44-77).

Extended metaphors included a metaphoric expression early on in the passage and elaborated on a topic using figurative language throughout the passage (e.g., the ‘zoo’ metaphor in the example given above). The target metaphor in each passage was what we call a ‘secondary metaphor’ that appeared towards the end of the passage (e.g., ‘protective gloves’ in the example above). While meaningful in the context, secondary metaphors were not directly related to the main topic of the passage. This was done in order to avoid intra-lexical priming from possible associates in the text.

We also constructed 21 single metaphors (average number of words: 57; range: 41-82) paired with the 21 extended metaphors. Single metaphors included a single metaphoric expression in the passage, which was the same as the secondary metaphors used in the extended metaphor condition. The context of the single metaphors was a literal version of the figurative context of extended metaphors. As in the case of the extended metaphors, we avoided intra-lexical priming of the target metaphor in the single metaphors.

In order to pre-test our materials, we constructed two different lists of materials each including the complete set of 42 metaphors but presented in reverse order (i.e. extended metaphors or single metaphors first). We administered the pre-test as a pen and paper questionnaire that included two rating tables underneath each item. Target metaphors were presented in bold in the passage and participants were asked to rate that particular metaphoric

expression on a 1-7 scale in terms of comprehensibility (1 = *Doesn't make any sense to me* – 7 = *Makes good sense to me*) and familiarity (1 = *Never heard it before* – 7 = *Common expression*).

Since the target metaphors were presented twice (once in each condition), participants were asked to base their familiarity ratings on their experience prior to the testing session in order to avoid artificially higher ratings in the second presentation. Also, participants were not allowed to look back in their questionnaires in order to avoid them giving the same score the second time they were presented with a metaphor. Participants were tested individually in a quiet room.

Results

For comprehensibility, the mean item ratings for the Single metaphor condition ranged from 4.3-6.2, with a global mean of 5.2 (SD 0.62). The mean item ratings for the Extended metaphor condition ranged from 4.7-6.6, with a global mean of 5.7 (SD 0.55). A paired t-test disclosed a significant difference between the ratings in the two conditions ($t = 3.91$, $df = 20$, $p < 0.001$).

For familiarity, the mean item ratings for the Single metaphor condition ranged from 1.8-5.2, with a global mean of 2.9 (SD 0.96). The mean item ratings for the Extended metaphor condition ranged from 1.9-4.7, with a global mean of 3.0 (SD 0.88). A paired t-test disclosed no significant difference between the ratings in the two conditions ($t = 1.38$, $df = 20$, $p = 0.18$).

Discussion

We conclude from the pre-testing of our materials that both the single and the extended metaphors were generally unfamiliar and highly comprehensible, notwithstanding the small but significant difference between comprehensibility ratings for the two conditions. We will further explore the possible consequences of this difference in subsequent analyses.

Experiment 1: Self-paced reading task

In this experiment, we contrast the processing of single metaphors, extended metaphors and literal expressions within a self-paced reading paradigm. As exemplified above, we control for variability in the reading times of different linguistic material by presenting the same critical phrases in the three conditions of interest. Within this paradigm, we would expect greater reading times where greater processing effort is required.

Our prediction, based on Carston's (2010) account, is that the reading times of the target expressions will be comparable when these are presented in the extended metaphor and in the literal context, as the linguistic material is initially interpreted purely literally in both cases. The reading times for the same material will be slower in the single metaphor condition, as in this case the figurative meaning is predicted to be immediately "cashed out" by the reader.

Method

Participants

84 undergraduate and postgraduate students from University College London took part in the experiment. They were all native speakers of English and participated for monetary compensation (£3).

Materials, design and procedure

We constructed 21 literal passages (average number of words: 59; range: 46-77) paired with the 21 extended and single metaphors. The literal passages included the same target expression (e.g., ‘protective gloves’ in the material sample given above) as the metaphors but used literally in an entirely literal context. The literal passages served as a control condition.

We constructed 3 lists of materials, each containing 7 items of each experimental condition (Extended Metaphor, Single Metaphor and Literal) plus 21 filler items and 2 warm-up items. Like the experimental items, 1/3 of the filler items were literal and 2/3 included figurative language. For the self-paced reading task, each passage was divided into 8 segments (average number of words: 7.4; range: 3-10). One of these segments was the target expression, which appeared in isolation (average number of words: 5.1; range: 2-9).

The experiment was programmed and run using E-prime. The experimental and filler items in each list were randomized individually for each participant.

Participants were randomly allocated to one of the 3 lists. They were asked to read the passages at a normal pace, making sure they understood each passage. Participants had to press the space bar to move the text forward. In between each passage, a row of asterisks appear on the screen. Participants were given a short memory test at the end of the self-paced reading task, in which they had to recognize 2 passages that they had previously read in a list of 4 passages. Participants were told about the memory test in the instructions in order to make sure that they read the passages carefully.

Results

All participants passed the memory test. Results were analysed from 83 participants (one set of results was discarded as the reading times were unrealistically short). The raw data comprise 581 reading times per condition (7 each from 83 participants). For these raw data,

the mean reading time for the critical segments in the Literal condition was 1457ms (SD 727ms), in the Extended Metaphor condition 1543ms (SD 764ms), and in the Single Metaphor condition 1578ms (SD 768ms).

To remove outliers, we discarded data points which were more than 2.5 standard deviations from the mean for each participant (thus discarding a total of 9 data points in the Literal condition, 19 in the Extended Metaphor condition and 15 in the Single Metaphor condition). For the resulting 1704 data points, the mean response times for the critical segments were 1421ms (SD 650ms) in the Literal condition, 1498ms (SD 716ms) in the Extended Metaphor condition, and 1537ms (SD 716ms) in the Single Metaphor condition.

The initial hypothesis under test was that there would be a difference between, on the one hand, Single items, and on the other, Extended and Literal items (considered as a single category). We tested for this difference by using a linear mixed effects model, using the `lmer` function from the `lme4` package (Bates et al. 2014) under the statistical computing language R (R Core Team 2014). In the model, we posited a main effect of Category (Single vs. Other) and random effects of Participant and Item, along with a random slope of Category by Participant (following Barr et al. 2013). The model disclosed a significant effect of Category, with the Single condition being read slower than the others (coefficient = 77.3, SE = 24.9, $t = 3.10$, $p < 0.01$).

However, follow-up pairwise comparisons of Condition levels, again using linear mixed models with random slopes, disclosed no significant difference between Extended and Single metaphors (coefficient = 38.4, SE = 29.7, $t = 1.30$, n.s.), but highly significant differences between each of these and the Literal condition (Literal vs. Extended: coefficient = 75.3, SE = 28.2, $t = 2.68$, $p < 0.01$; Literal vs. Single: coefficient = 114.7, SE = 28.6, $t = 4.00$, $p < 0.001$).

We also wished to explore the relationship between the reading times for metaphoric expressions and their judged comprehensibility, as elicited in the pre-test. For the Single condition, we ran a further linear mixed model, positing a main effect of Comprehensibility, a random effect of Participant, and a random slope of Comprehensibility by Participant. There was a significant effect (coefficient = -121.8, SE = 42.1, $t = -2.89$, $p < 0.01$), suggesting that more comprehensible single metaphors were read faster. We did the same for the Extended condition, and a similar numerical effect did not reach significance (coefficient = -76.5, SE = 43.4, $t = -1.76$, $p = 0.084$).

Discussion

The first comparison between reading times for the target expressions (e.g., ‘without protective gloves’) in the single-metaphorical condition, on the one hand, and in the extended metaphorical and literal conditions, on the other, supported the distinction predicted by Carston’s account, with reading times being significantly longer in the single metaphor condition. However, further pairwise comparisons seemed to support the standard RT account, with response times being significantly faster in the literal condition than in the two metaphorical conditions. The lack of a significant difference between the two metaphorical conditions also suggests that their difference in comprehensibility is very limited in its effect.

Self-paced reading times, however, might not be sensitive enough to reveal nuanced differences in the processing of single and extended metaphors. It is for this reason that in Experiment 2 we used the same materials in a natural reading task coupled with eye-tracking. Relative to self-paced reading, eye-tracking during natural reading should reveal finer-grained differences between the processing of the target expressions in the different contexts as it allows making a distinction between early and late processing measures.

Experiment 2: Eye-tracking during natural reading

We implemented an eye-tracking experiment to provide a more precise measure of reading times over the same materials. In particular, we wished to explore whether the same distinctions between conditions observed in Experiment 1 were evident at all stages of processing, or a more complex pattern emerged when distinguishing early and late processing measures. If Carston (2010) is correct, the early processing measures should yield comparable results in the extended metaphor and literal conditions (since the target expressions would initially be processed literally in both conditions) but distinct results in the single metaphor condition (in which early processing already involves computation of the figurative meaning). By contrast, if we further assume, in line with Carston (2010), that readers of extended metaphors start to compute their figurative meanings in later processing, we would expect late processing measures to yield comparable results in the two metaphor conditions, both distinct from the literal condition (in which this additional processing is not necessary at any stage). This is a reasonable assumption considering our materials: the target expressions always appear in the second half of the extended metaphorical passages, by which stage readers should already have ‘switched’ to the literal-imagistic route of metaphor interpretation.¹ As in Experiment 1, none of these predictions follow from the standard Relevance-Theoretic account of loose use.

Method

¹ It would also be compatible with Carston’s (2010) account if deeper and more reflective interpretation processes were to take place towards the end of figurative passages, but we did not test such off-line interpretation processes in our study.

Participants

61 native English speakers were tested. They were recruited via University College London Psychology Subject Pool, and received £4 for their participation. Data from 11 participants were excluded due to failed calibration or track loss.

Materials and Procedure

The materials from the previous experiment were reused. Each participant read 21 experimental sentences, 7 in the Extended Metaphor condition, 7 in the Single Metaphor condition, and 7 in the Literal Control condition. The sentences were presented in a random order.

The experiment was programmed and run using the EyeTrack software version 7.10m (see <http://www.psych.umass.edu/eyelab>). Eye movements were recorded using the Eyelink 1000 eye tracker developed by SR Research Ltd. The eye tracker was in a desktop mount configuration, and the recording was sampled at 1000Hz from the right eye. A chin and forehead rest was used to stabilize the participants' head position. Viewing distance was 60cm. An experimenter sat on a nearby table facing the side of the participants. The experimenter and the participant could not see each other's screens, but the experimenter could view the participant's eye movements on her system. The screen resolution was set to be 1024 x 768 pixels. Texts were presented in size 14 Arial font, starting from 100 pixels from the left and 100 pixels from the top. The sentences wrap around at about 600 pixels from the left. At the start of the experiment, participants were briefed and instructed to read passages silently at their normal reading speed. Participants' eye movements were calibrated and then validated using nine fixation points; when the calibration was classified as "good" by the EyeLink software, the experiment proceeded.

In the experiment, participants were first presented with three practice passages, of similar lengths to the experimental items but containing no metaphors. The practice was followed by 21 experimental passages. Before each item, a fixation cross appeared in the centre of the screen. If the fixation was accurate, the experimenter would press a key to release the next item. The entire session lasted around 20 minutes per participant.

Data analysis

Individual participants' eyetracking data was checked and preprocessed using the EyeDoctor software (see <http://www.psych.umass.edu/eyelab>). During this process, blinks and track loss were checked. If a blink or track loss occurred in the critical regions (corresponding with the target expressions), the trial was removed. This resulted in 9% data loss. Reading behaviours of the critical region and the word immediately after the critical region were analysed. We calculated seven processing measures, defined as follows.

First fixation duration: the duration of initial fixation(s) in a region. We calculated the sum of all initial fixations in the critical region (if there was more than one initial fixation) rather than just the very first fixation because the critical region of many items contains several words, likely to incur more than one fixation. Thus, the very first fixation alone would not be an informative measure.

First pass time: the duration of all fixations in a region before first moving out of the region (this includes regressions back to earlier parts of the sentence before moving on).

Second pass time: the sum of durations of fixations in a region in subsequent passes after first moving out of the region.

Spillover first fixation duration: the duration of the very first fixation after leaving the region.

Total time: the sum of the durations of all fixations in a region.

Regressions in: the proportion of trials on which a regressive eye movement was made into a region.

Regressions out: the proportion of trials on which a regressive eye movement was made out of a region.

Following the consensus in the literature (e.g., Rayner 1998, Reichle et al. 1998), we assume that first fixation duration and first pass time reflect initial processing, while second pass time, proportions of regressions in and out and spillover first fixation duration reflect later processes. We assume that the total time reflects the overall processing effort.

Note that spillover first fixation duration is not only affected by the processing of the critical region, but also by the length and frequency of the word immediately following the critical region. This word is identical in the single and extended metaphor conditions, but is sometimes different for the literal condition (see sample item above). To ensure that this measurement would be informative about the processing of the critical region, we compared the length of the words immediately following the critical region as well as their frequency, extracted from the British National Corpus (Leech et al. 2001). The average length is 3.86 letters (SD 1.96) for the single and extended metaphor conditions and 4.05 letters (SD 1.94) for the literal condition. The average frequency is 1.72% (SD 2.6%) for the single and extended metaphor conditions and 1.28% (SD 2.1%) for the literal condition. Neither the lengths ($t(20) = -0.45, p = 0.66$) nor the frequencies ($t(19) = 1.25, p = 0.23$) differ significantly. If anything, the first word in the literal condition is numerically longer and less frequent than that of the metaphor conditions, both suggesting potential longer durations for the literal condition (Rayner 1998).

Results

Analyses were conducted by fitting a linear (duration measurements) or logistic (proportion measurements) mixed effect model for each measurement, using the lmer function from the lme4 package (Bates et al. 2014) under the statistical computing language R (R Core Team 2014). The fixed effect was Condition, and the random effects were Participants and Items. For all analyses, we report the estimates, standard errors, *t*-values or *z*-values as well as the *p*-values for the fixed effects. For the three level analysis (Single Metaphor, Extended Metaphor and Literal), we also report the *F*-values and *p*-values for the main effect of Condition.

Two-level analysis: first we compared the Single Metaphor condition with the other two conditions, Extended Metaphor and Literal (referred to as the “Literals” in Table 1), which should be comparable according to Carston (2010). Following the recommendations of Barr et al. (2013), we tried to fit maximal models which included random slopes and intercepts for both subjects and items. When fitting maximal models ran into convergence errors, we would modify the model by assuming independent intercepts and slopes for the random effects. When this ran into convergence errors, we would remove the random slopes for the random effects and keep the random intercepts. For this analysis, models for first pass time, second pass time and total time were maximal, models for first fixation duration assumed independent intercepts and slopes for the random effects, and models for spillover first fixation, proportions of regressions in and regressions out had random intercepts only.

Table 1 reports the sample means of duration of proportion measures and model fitting results for the two conditions. The averages of all five duration measurements of the Single Metaphor condition were longer than the Literal conditions. Among them, the differences in total time and spillover first fixation duration were significant, and the difference in first

fixation duration was borderline significant ($p = 0.05$). This suggests that the critical region requires more processing effort in the Single Metaphor condition than in the other two conditions, and this difference is incurred from the earliest processing stages. There were no significant differences in regressions in and out.

Table 1 Sample means and mixed effects regression results – two levels (Single versus Literals)

	Sample Mean		Model estimates		Standard Errors		<i>t</i>	<i>p</i>
	Single	Literals	Single	Literals	Single	Literals		
First fixation (ms)	242	230	243	229	8.02	6.89	2.05	0.05
First pass time (ms)	491	467	468	448	52.88	24.17	0.84	0.41
Second pass time (ms)	558	432	486	403	85.22	65.38	1.27	0.22
Total time (ms)	634	575	620	558	73.5	26.85	2.3	0.03
Spillover first fixation (ms)	245	230	248	233	8.99	7.49	2.05	0.04
							<i>z</i>	<i>p</i>
Regressions in (%)	9%	11%	8%	11%	0.31	0.33	-1.4	0.16
Regressions out (%)	23%	23%	20%	20%	0.21	0.17	-0.09	0.93

Three-level pairwise analysis: we then separated the Extended Metaphor condition from the Literal condition, and compared three levels: Single Metaphor, Extended Metaphor, and Literal. The models for first pass time and second pass time were maximal. The models for first fixation and spillover first fixation assumed independent intercepts and slopes for the random effects. The models for total time, regressions in and regressions out contained random intercepts only.

The results are reported in Table 2 and Table 3. Numerically, the Single Metaphor condition has longer mean durations in all five duration measures than the Extended Metaphor and the Literal conditions. Condition has a significant main effect on total time (Table 2). There is a significant difference between Single Metaphor and Literal, and a trending difference between Single Metaphor and Extended Metaphor. There is also a significant difference between Extended Metaphor and Literal (Table 3). In addition, pairwise comparison shows a trending difference ($p = 0.05$) between Single and Extended Metaphor conditions for first fixation duration (Table 3).

Table 2 Sample means and mixed effects regression results – three levels (Single Metaphor, Extended Metaphor & Literal)

	Sample mean			Model Estimates			Standard Errors				
	Single	Extended	Literal	Single	Extended	Literal	Single	Extended	Literal	<i>F</i>	<i>p</i>
First fixation (ms)	242	229	230	243	229	229	7.45	7.56	0.93	2.21	0.12
First pass time (ms)	491	478	454	468	460	437	20.13	24.93	20.63	0.48	0.63
Second pass time (ms)	558	412	461	486	401	390	70.02	40.49	60.93	0.96	0.39
Total time (ms)	635	598	553	623	582	531	23.22	63.98	23.26	7.51	6E-04
Spillover first fixation (ms)	245	234	227	249	239	228	10.37	11.88	11.77	1.87	0.17
Regressions in (%)	9%	15%	8%	11%	14%	11%	0.35	0.25	0.25	2.37	0.09
Regressions out (%)	23%	23%	24%	20%	19%	20%	0.2	0.21	0.19	0.07	0.94

Table 3 Pairwise *t* values and *p* values from mixed effects regression

	Single – Literal		Single - Extended		Extended – Literal	
	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>

First fixation (ms)	1.58	0.12	1.96	0.05	-0.12	0.9
First pass time (ms)	0.99	0.33	0.28	0.78	0.67	0.51
Second pass time (ms)	1.39	0.17	1.22	0.23	0.18	0.86
Total time (ms)	3.00	0.007	1.68	0.09	2.2	0.03
Spillover first fixation (ms)	1.9	0.06	0.92	0.37	0.9	0.38
	<i>z</i>	<i>p</i>	<i>z</i>	<i>p</i>	<i>z</i>	<i>p</i>
Regressions in (%)	0.12	0.9	-1.75	0.08	-1.9	0.06
Regressions out (%)	-0.07	0.95	0.25	0.8	0.35	0.73

Discussion

The initial two-level analysis once again suggests a difference along the lines posited by Carston (2010) between, on the one hand, single metaphors, and on the other, extended metaphors and literal items, which is particularly manifest in total reading time. However, once again, the follow-up three-level analysis disclosed a more complex picture. First fixation time, regarded as a measure of early processing, appears to exhibit the distinction posited by Carston (2010), with the single condition giving rise to longer reading times, and no difference evident between the extended and literal conditions.

For measures that are assumed to be reflective of later processing, the single condition again gives rise to numerically longer reading times, but the differences from the extended condition are no longer significant, suggesting that the two metaphoric conditions are beginning to pattern together more closely in later processing. These results bear out the predictions discussed above, and we interpret them as supportive of the second processing route posited by Carston (2010). In this processing mode, readers would have first extracted

the literal meanings of the extended metaphors and then started to compute their figurative meanings in later processing.

Experiment 3: Cued recall

In the last experiment of the study, we wanted to explore what effect potential differences in processing might have on readers' recall of single and extended metaphors (for earlier studies of metaphor comprehension using cued recall, see McGlone (1996), Katz & Law (2010)). If single metaphors involve local pragmatic adjustment of the lexical meaning of the metaphor vehicle, but extended metaphors do not (because their literal meaning is processed globally for its metaphorical meaning), then the target expressions in our materials should be better remembered in the single-metaphorical condition than in the extended metaphorical condition. In this sense, the target expressions would have higher 'distinctiveness' in the former condition than in the latter; that is, they would stand out more clearly in the single-metaphorical contexts, affecting memory positively (Hunt 1995, Hunt & Worthen 2006).

It is worth stressing that the target expressions in our materials were not associated to previous words in the context in either metaphorical condition (in the extended metaphor condition, in particular, the target expressions were 'secondary metaphors'). Therefore, it was not simply the case that the target expressions were less distinctive in the extended metaphorical condition because they were more strongly primed by other associates in the figurative passage. Our prediction is therefore based only on potential differences in processing between single and extended metaphors, as predicted by Carston (2010), and the effect that such processing differences would have on the distinctiveness of the two metaphor types. Note, in this respect, that the prediction following from Carston's (2010) account regarding the memorability of single and extended metaphors is that this is related to, and not

an alternative explanation for, their potentially different distinctiveness (Hunt & Worthen, 2006). In contrast, according to the standard Relevance-Theoretic account of metaphor interpretation, both single and extended metaphors would be processed via the same pragmatic mechanisms, not necessarily resulting in any differences in distinctiveness and memorability.

Method

Participants

28 undergraduate and postgraduate students from University College London took part in the experiment. They were all native speakers of English and participated for monetary compensation (£4).

Materials, design and procedure

We constructed two booklets including 40 metaphors from our materials: 20 extended and 20 single counterbalanced by booklet. For the memory prompts we used a strong associate of the metaphor vehicles (e.g., ‘without protective gloves’ > FINGERS). The structure of the booklet was as follows: on the first page of each item, participants were asked to read a passage carefully but only once. At the bottom of the page there was a table with 5 cells, and participants were asked to fill in 4 cells counting backwards in 3s from the number on the first cell (e.g., **73** – 70 – 67 – 64 – 61). The point of this secondary task was to prevent participants from responding to the cued-recall test from short-term memory. On the other side of the page, participants were asked to write down the first word that came to mind when reading a prompt (e.g., FINGERS: “Gloves”). Participants were not allowed to turn back the page to look at the passage, but they were allowed to complete the memory test in their own time. Participants were randomly allocated to one of the two booklets.

Results

We compared the rates of recall for each participant between Extended (mean 5.2, SD 2.5) and Single (mean 6.5, SD 2.3) conditions using a paired t-test. This disclosed superior recall in the Single condition ($t = 3.10$, $df = 27$, $p < 0.01$).

To test whether there were effects of comprehensibility on recall, we examined the total scores by item, and compared these with the comprehensibility ratings in the relevant condition (Extended or Single, as appropriate). For the Extended metaphors, there was a moderate positive correlation between comprehensibility ratings and recall (Pearson's $r = 0.304$). For the Single metaphors, there was also a moderate positive correlation between comprehensibility ratings and recall (Pearson's $r = 0.32$).

Discussion

The results of Experiment 3 supported the hypothesis outlined above, with the target expressions being better recalled in the single-metaphorical condition (in which they would be interpreted by a local process of meaning modulation) than in the extended metaphorical condition (in which they would be interpreted literally, in line with the rest of the figurative passage). This appears to cohere better with an account in which single and extended metaphors are processed in different ways (such as that of Carston (2010)) than one in which metaphors of these different types are treated identically by the reader/hearer (such as the standard RT account of metaphor in terms of ad hoc concepts, e.g., Sperber & Wilson 2008).

General discussion

This study investigated the distinction between the processing of single and extended metaphors proposed by Carston (2010) in her recent account of metaphor interpretation. In line with the standard Relevance-Theoretic account of metaphor as loose use, Carston argues that single metaphors are interpreted by a local pragmatic process of meaning adjustment, resulting in the construction of an ad hoc concept. In extended metaphorical passages, by contrast, the reader switches to a second processing mode, due to the mutual priming and reinforcing of the literal meaning by the various semantic associates in the passage. According to Carston, this second processing mode is characterised by a slower, more reflective type of processing in which the literal meaning of the whole passage is metarepresented and the intended figurative implications are only derived at a later stage of processing.

The results of our study offer initial support to the view that single and extended metaphors are processed differently, in line with Carston's (2010) account of metaphor interpretation. The results of Experiment 1 revealed longer reading times for the target expressions in the single metaphor condition compared to the extended metaphor and literal conditions (both of which would be interpreted literally, at least initially). Further analyses, however, revealed that this difference was mainly due to the faster reading times observed in the Literal condition. Experiment 2 offered more nuanced results with the use of eye-tracking during natural reading: early processing measures revealed comparable results in the extended metaphor and literal conditions, and suggested a small but significant difference between the extended and single metaphor conditions, in the direction predicted by Carston (2010). However, the distinction between the extended and single metaphor conditions ceased to be significant in later processing measures, again in accordance with the predictions. We interpreted this pattern of results as suggestive that readers had started to derive the figurative meanings of the extended metaphorical passages in later processing.

Finally, Experiment 3 tested the hypothesis that the same expression would be more distinctive and therefore better recalled as a single metaphor in an otherwise literal passage, than as a figurative expression in an extended metaphorical passage. Such a difference in cued recall would be consistent with Carston's distinction between the processing of single and extended metaphors, while it would not be necessarily predicted by the standard RT account. The results of the third experiment supported the experimental hypothesis and revealed that our target expressions were better recalled in the single metaphor condition than in the extended metaphor condition.

Taken together, the processing differences observed in our experiments suggest that there is validity to the claim that extended metaphors have the potential to engage a different mode of processing to that customarily engaged by single metaphors. However, one might criticize that by manipulating the contexts preceding the target expressions, we do not only distinguish the different conditions but also introduce possible other factors that could in principle affect the processing of the target expressions (e.g., varying word lengths and frequencies, different syntactic constructions, potential priming effects, and so on)². Most of these factors should not result in any specific hypotheses that could be pinned against the two Relevance-

² While we acknowledge this shortcoming of our paradigm, we see it as inevitable insofar as it would be almost impossible to control for such a vast suite of potential factors in the construction of the materials. In other words, it would be extremely difficult to construct a battery of single metaphors, extended metaphors and literal controls that were preceded by contexts containing words of comparable length and frequency, identical syntactic constructions, the same potential priming effects, etc., while making sure that the passages are highly comprehensible.

Theoretic accounts of metaphor interpretation investigated in this study. It is conceivable that the flow of the narrative is more uneven in the simple condition compared to the extended condition, which could be associated with different processing costs, but none of the accounts offered so far make clear predictions about how this would influence our outcome measures.

In general, given the limited scope of this study, it would be a considerable overreach to assume that the mode of processing depends wholly upon whether the metaphor is single or extended: in fact, if we grant the existence of these two modes, the precise conditions under which each is engaged remains a matter for further study (see Carston (2010: 311) for a discussion of metaphors that are not particularly lengthy but may nonetheless be interpreted via the secondary route). Future theoretical and experimental work should therefore establish which factors might affect the use of one or the other processing mode (e.g., extension of the metaphor, topic of the discourse, imagistic component, etc.).

The aim of the present study was to tease apart the standard RT account of loose use from the more recent proposal put forward by Carston (2010). While supportive of the latter, the processing differences observed in our study are generally compatible with other approaches to cognition (e.g., Clark, 2013). Pickering and Garrod (2013), for example, propose a theory that integrates language production and comprehension on the basis of prediction. It follows from this model that the interpretation of an extended metaphor would rely on the various semantic associates in the text priming each other and creating the expectation that the passage will continue in a coherent manner. While this much is compatible with Carston's second processing route, a potential difference arises when dealing with the type of secondary metaphors that we used in our study: while interpretable in the passage, our target metaphors were not directly primed by other semantic associates earlier in the text. Therefore, Pickering

and Garrod's model would predict that the interpretation of secondary metaphors in extended figurative passages should be more effortful than that of primary metaphors (which are semantically related to the main topic of the passage). By contrast, such a prediction would not necessarily follow from Carston's model because the readers of our materials should have already switched to the literal/imagistic interpretive route when encountering the secondary metaphors, and this processing mode relies not so much on priming but on meta-representing the entire passage as an imagined world.

Future studies might aim to provide more direct evidence for Carston's (2010) account of two processing modes in metaphor interpretation. A promising line of investigation would be to compare the interpretation of the metaphors appearing at the beginning and the end of an extended figurative passage. In particular, it would be relevant to consider whether the literal meaning of these expressions is less accessible at the beginning of the text – when the reader is, according to Carston's model, still trying to modulate the meaning of each metaphor individually – than at the end of the passage, when the reader should have already switched to the second processing route.

Other extended figurative uses

Another open question is whether the literal-imagistic interpretation that characterises the reading of extended metaphors is the same type of interpretation that arises in the case of allegories, parables and the like. Carston and Wearing (2011) have recently pointed out two important similarities between allegories and extended metaphors in relation to the second mode of processing proposed by Carston (2010): first, both literary devices are developed over the length of a whole passage or narrative, and second, they both have a superficial, literal meaning and a deeper, figurative meaning. On the basis of these similarities, Carston

and Wearing argue that it is likely that the interpretation of both analogies and extended metaphors includes a phase in which the literal meaning of the passage is metarepresented and parallel, deeper meanings are derived in further interpretive processes. On the other hand, the interpretation processes by which allegories and extended metaphors are fully understood are unlikely to be identical, given that allegories, unlike extended metaphors, are fully coherent at the literal level and can be superficially interpreted at that level alone (see Carston and Wearing (2011: 307-309) for further discussion and examples, and Gibbs and Blackwell (2012) for a case study of allegory interpretation).

In conclusion, further theoretical work needs to be done in mapping out the processes underlying the interpretation of single and extended metaphors (a distinction currently only drawn by Carston (2010)), as well as allegories, parables and other extended figurative uses. Moreover, the possibility of categorical differences in the mode of metaphor processing suggested by our results urgently calls for further experimental research across this field, not least to ascertain whether existing findings about the time-course of interpretation can safely be generalised to all categories of metaphoric expression.

Acknowledgements

This research was supported by a Research Project Grant from the Leverhulme Foundation ('Understanding Metaphor: *Ad Hoc* Concepts and Imagined Worlds'; Ref: F/07 134/DP) awarded to the first author as Research Associate in the project. The authors would like to thank Robyn Carston (PI in the project) for her helpful comments on earlier drafts of the paper.

References

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255-278.
- Bates, D., Maechler, M., Bolker, B. & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R package version 1.0-6*. <http://CRAN.R-project.org/package=lme4>
- Carston, R. (2010). XIII—Metaphor: Ad Hoc Concepts, Literal Meaning and Mental Images. In *Proceedings of the Aristotelian Society*, 110, 295-321. Blackwell Publishing Ltd.
- Carston, R., & Wearing, C. (2011). Metaphor, hyperbole and simile: A pragmatic approach. *Language and Cognition*, 3, 283-312.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36, 181-204.
- Davidson, D. (1978/1984). What Metaphors Mean. *Critical Inquiry*, 5, 31–47. Reprinted in his *Inquiries into Truth and Interpretation*, 1984, pp. 245–64. Oxford: Oxford University Press.
- Gentner, D., Imai, M., & Boroditsky, L. (2002). As time goes by: Evidence for two systems in processing space→ time metaphors. *Language and Cognitive Processes*, 17, 537-565.
- Gibbs, R. W. & Blackwell, N. (2012). Climbing the ladder to literary Heaven: A case study of allegorical interpretation of fiction. *Scientific Study of Literature*, 2, 199-217.
- Glucksberg, S. & Haught, C. (2006). On the relation between metaphor and simile: When comparison fails. *Mind & Language*, 21, 360-378.
- Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review*, 97, 3.

- Glucksberg, S., Newsome, M. R., & Goldvarg, Y. (2001). Inhibition of the literal: Filtering metaphor-irrelevant information during metaphor comprehension. *Metaphor and Symbol, 16*, 277-298.
- Grice, H. P. (1975). Logic and Conversation. In P. Cole and J. Morgan (Eds.), *Syntax and Semantics, 3: Speech Acts*, pp. 41–58. New York: Academic Press.
- Hunt, R. R. (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review, 2*, 105-112.
- Hunt, R., & Worthen, J. B. (2006). *Distinctiveness and memory*. Oxford University Press.
- Katz, A., & Law, A. (2010). Experimental support for conceptual metaphors with an episodic memory task. *Metaphor and Symbol, 25*, 263-270.
- Keysar, B., Shen, Y., Glucksberg, S., & Horton, W. S. (2000). Conventional language: How metaphorical is it?. *Journal of Memory and Language, 43*, 576-593.
- Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science, 4*, 195-208.
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: based on the British National Corpus*. Longman.
- McGlone, M. S. (1996). Conceptual metaphors and figurative language interpretation: Food for thought? *Journal of Memory and Language, 35*, 544-565.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*, 329-347.
- R Core Team (2014). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. <http://www.R-project.org/>.

- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- Recanati, F. (2004). *Literal Meaning*. Cambridge: Cambridge University Press.
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105, 125.
- Rubio-Fernández, P. (2007). Suppression in metaphor interpretation: Differences between meaning selection and meaning construction. *Journal of Semantics*, 24, 345–71.
- Searle, J. R. (1979). Metaphor. In his *Expression and Meaning*. Cambridge: Cambridge University Press.
- Shen, Y., & Balaban, N. (1999). Metaphorical (in) coherence in discourse. *Discourse Processes*, 28, 139-153.
- Sperber, D. & Wilson, D. (2008). A deflationary account of metaphors. In: R. W. Gibbs (ed.), *Cambridge handbook of metaphor and thought*, pp. 84-105. Cambridge: Cambridge University Press.
- Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behaviour*, 18, 645-659.
- Thibodeau, P., & Durgin, F. H. (2008). Productive figurative communication: Conventional metaphors facilitate the comprehension of related novel metaphors. *Journal of Memory and Language*, 58, 521-540.
- Wilson, D. & Carston, R. (2007). A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts. In: N. Burton-Roberts (ed.) *Pragmatics*, pp. 230-260. Basingstoke: Palgrave Macmillan.

Wilson, D., & Sperber, D. (2002). Truthfulness and relevance. *Mind*, 111, 583-632.

Appendix: Sample items from the experimental materials including extended metaphors (EM), single metaphors (SM) and literal controls (LC) with the target expression in bold.

1/EM: In Martha's experience, depression was a toad that squatted wetly on your head until it finally gathered the energy to slither off. Yet the unhappiness she had been living with for the last ten days was quite different. It was frantic and aggressive. **It had sharp claws.** It didn't sit, it assailed.

1/SM: Martha knew what depression was like. She often had bad periods when she didn't want to do anything and had problems getting out of bed. Yet the unhappiness she had been experiencing for the last ten days was quite different. **It had sharp claws.** It was clear Martha needed professional help this time.

1/LC: Martha regretted having brought home the stray cat that she had found in the park. At times it was frantic and aggressive. In the last two days, it had torn the curtains in the living room and destroyed the four legs of the kitchen table. **It had sharp claws.** Martha was afraid it might be too late to adopt such an old cat.

2/EM: John's ambition was a racing car. He took every opportunity he had to overtake his rivals and challenge his teammates. He tore round and round the circuit until **the grandstand would give him a standing ovation.** John's boss praised him for his victories but secretly wondered where the finish line would be.

2/SM: John was overly ambitious. Everything he did at work was part of a strategy to get a new promotion. He was very competitive and always expected that **the grandstand would**

give him a standing ovation. John's boss was pleased with John's work, but his colleagues were tired of his tactics.

2/LC: John was training to become a Formula 1 driver That was his only dream and he was determined to work hard until the day he could tear round and round the circuit and **the grandstand would give him a standing ovation.** But for now it was all about working hard.

3/EM: Helen spoke with the soothing tones of a skilled healer. If she was in the right mood, her words had a soft edge that could magically cure you of all your ailments without any medicine. But if she thought a treatment was needed, **she would apply her scalpel close to your heart,** so you might find yourself wounded by the long blade of a 'why' or the incurable silence of a full stop.

3/SM: Helen was very articulate. She always explained everything clearly and in just enough detail to make it interesting. If you needed help, she could make you regain perspective with just a few words. But her words could also be acutely penetrating. Sometimes **she would apply her scalpel close to your heart,** and her words would leave you hurting for days.

3/LC: Helen had spent eight years in medical school and another four years training as a cardiac surgeon before she got her first job. But she never questioned her choices. She had wanted to become a doctor ever since she was a kid. Helen felt most empowered at the start of a surgery, when **she would apply her scalpel close to the heart,** her patient's life in her hands.

4/EM: When Gina's employees were cooperative, they could achieve great things together. They were a choir in perfect harmony, the delicate notes of the sopranos blending perfectly with the powerful rumble of the basses. After the performance, Gina had merely to pick up **the bouquets thrown onto the stage.** But when they were uncooperative, the resulting cacophony assailed the senses.

4/SM: When Gina's employees were cooperative, they could achieve great things together. They complemented one another beautifully, the imagination of the creative thinkers being made real by the labour of the diligent workers. On these occasions, Gina had merely to pick up **the bouquets thrown onto the stage**. But when they were uncooperative, the company rapidly descended into chaos.

4/LC: Gina's job with the opera company could be difficult. She was responsible for making sure the performances went smoothly, which was no easy task when the singers were at their most egotistical. When things had gone well, Gina had merely to pick up **the bouquets thrown onto the stage**. On other occasions, she would spend hours trying to placate the furious performers.

5/EM: Louise had remarkable mental abilities. Her brain was a computer, in which every data file was neatly stored away in the appropriate directory and could be accessed and operated upon at the touch of a button. However, just lately, **there was the occasional power outage**. Abruptly she would find herself unable to respond to the simplest request.

5/SM: Louise had remarkable mental abilities. She possessed formidable command of facts, and could supply information and reason about it at impressive speed. However, just lately, **there was the occasional power outage**. Abruptly she would find herself unable to respond to the simplest request.

5/LC: The executives at the company were very concerned about the electricity supply. All the data on the network was stored on servers which had to be constantly running. However, just lately, **there was the occasional power outage**. Today they were considering whether to invest in backup generators.